ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10

DOI: http://doi.org/10.5281/zenodo.3662434

Applications of Finite Markov Chains to Artificial Intelligence

Michael Gr. Voskoglou

Department of Mathematical Sciences, School of Technological Applications,
Graduate Technological Educational Sciences Institute of Western Greece, Patras, Greece
voskoglou@teiwest.gr ; mvoskoglou@gmail.com

ABSTRACT:
The theory of MCs is a smart combination of Linear Algebra and Probability theory offering ideal conditions for the study and mathematical modelling of situations depending on random variables and finding important applications to problems of Artificial Intelligence. In the paper at hands an absorbing Markov chain is introduced on the phases of decision making and an application is presented illustrating the importance of the constructed model in practice. Further, the Case-Based Reasoning process is modeled with the help of an ergodic Markov chain defined on its steps and through it a measure is obtained for the effectiveness of a Case-Based Reasoning system.
Keywords: Markov Chains (MC's), Absorbing MC's (AMC's), Ergodic MC's (EMC's), Artificial Intelligence (AI), Decision Making (DM), Case-Based Reasoning (CBR).

INTRODUCTION

Artificial intelligence (AI) is the branch of Computer Science that focuses on the creation of intelligent machines which work and react like humans. The term AI **was** first coined by John McCarthy (Figure 1) in 1956 when he held the first academic conference on the subject. But the journey to understand if machines can truly think began much before that.

Figure 1: J. McCarthy (1927-2011)

AI has roots in mathematics, engineering, technology and science and as a synthesis of ideas from all those fields has created a new situation that is only just beginning to generate enormous changes and benefits to the human society.

Probability theory, dealing with situations of uncertainty caused by randomness, is included among the mathematical tools used in AI applications. In particular, the *Markov Chain (MC)* theory is a smart combination of Probability and Linear Algebra that is used in problems of AI to model something that is in discrete states, but it is not fully understood how it is evolved [1].

The purpose of the paper at hands is to present applications of finite MC's to problems of AI involving *Decision Making (DM)* and *Case-Based Reasoning (CBR)* systems. The paper is organized as follows: In the second section the elements of theory of finite MC's are recalled which are necessary for the understanding of the rest of the work. In the third section an *absorbing MC (AMC)* is introduced on the phases of the decision making and an application is presented illustrating the importance of the constructed model. In the fourth section the CBR process is modeled with the help of an *ergodic MC (EMC)* having as states the steps of CBR and through it a measure is obtained for the effectiveness of a CBR system. The paper ends with the final conclusions presented in the fifth section.

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10

DOI: http://doi.org/10.5281/zenodo.3662434

MARKOV CHAINS

Roughly speaking a *Markov chain (MC)* is a stochastic process that moves in a sequence of steps (phases) through a set of states and has a *one-step memory.* In other words, the probability of entering a certain state in a certain step depends on the state occupied in the previous step and not in earlier steps. This is known as the *Markov property*. However, for being able to model as many real life situations as possible by using MCs, one could accept in practice that the probability of entering a certain state in a certain step, although it may not be completely independent of previous steps, it mainly depends on the state occupied in the previous step [2].

The basic concepts of MCs were introduced by Andrey Markov (Figure 2) in 1907 on coding literal texts.

Figure 2: A. Markov (1856-1922)

Since then the MC theory was developed by a number of leading mathematicians, such as A. Kolmogorov, W. Feller, etc. However, only from the 1960's the importance of this theory to the natural, social and applied sciences has been recognized [1-7].

1. Finite Markov Chains

When the set of states of a MC is a finite set, then we speak about a *finite MC.* For general facts on finite MCs we refer to Chapter 2 of the book [8].

Let us consider a finite MC with n states, say S_1, S_2, ..., S_n, where n is a non negative integer, $n \geq 2$. Denote by p_{ij} the *transition probability* from state S_i to state S_j , i, j = 1, 2,...., n ; then the matrix $A = [p_{ij}]$ is called the *transition matrix* of the MC. Since the transition from a state to anyone of the other states (including its self) is the certain event, we have that

$$p_{i1} + p_{i2} + \ldots + p_{in} = 1(1), \text{ for } i = 1, \ldots, n$$

(1)

The row-matrix $P_k = [p_1^{(k)}\ p_2^{(k)} \ldots\ p_n^{(k)}]$, known as the *probability vector* of the MC, gives the probabilities $p_i^{(k)}$ for the MC to be in state i at step k , for i = 1, 2,...., n and k = 0, 1, 2,.... Obviously we have again that

$$p_1^{(k)} + p_2^{(k)} + \ldots + p_n^{(k)} = 1 \quad (2)$$

Using conditional probabilities on can show ([8], Chapter 2, Proposition 1) that $P_{k+1} = P_k A$ (3), for all non negative integers k. Therefore a straightforward induction on k gives that $P_k = P_0 A^k$ (4), for all integers $k \geq 1$. Equations (3) and (4) enable one to make *short run* forecasts for the evolution of the various situations that can be represented by a finite MC. In practical applications we usually distinguish between two types of finite MCs, the AMCs and the EMCs.

2. Absorbing Markov Chains

A state of a MC is called *absorbing* if, once entered, it cannot be left. Further a MC is said to be an AMC if it has at least one absorbing state and if from every state it is possible to reach an absorbing state, not necessarily in one step.

Working with an AMC with k absorbing states, $1 \leq k < n$, one brings its transition matrix A to its *canonical* (or *standard*) *form* A^* by listing the absorbing states first and then makes a partition of A^* to sub-matrices as follows

$$A^* = \begin{bmatrix} I_k & | & O \\ - & | & - \\ R & | & Q \end{bmatrix} \qquad (5).$$

In the above partition of A^*, I_k denotes the unitary k X k matrix, O is a zero matrix, R is the *(n – k)* X k transition matrix from the non-absorbing to the absorbing states and Q is the *(n – k)* X *(n – k)* transition matrix between the non absorbing states.

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10

DOI: http://doi.org/10.5281/zenodo.3662434

It can be shown ([9], Section 2) that the square matrix I_{n-k} - Q, where I_{n-k} denotes the unitary n-k X n-k matrix is always an invertible matrix. The *fundamental matrix N* of the AMC is defined to be the inverse matrix of $I_{n-k} - Q$. Therefore ([10], Section 2.4)

$$N = [n_{ij}] = (I_{n-k} - Q)^{-1}$$

$$= \frac{1}{D\ (I_{n-k} - Q)} adj\ (I_{n-k} - Q) \qquad (6).$$

In equation (6) $D(I_{n-k} - Q)$ and $adj\ (I_{n-k} - Q)$ denote the determinant and the *adjoin* of the matrix $I_{n-k} - Q$ respectively It is recalled that the adjoin of a matrix M is the matrix of the *algebraic complements* of the *transpose* matrix M^t of M, which is obtained by turning the rows of M to columns and vice versa. It is also recalled that the algebraic complement m_{ij}' of an element m_{ij} of M is calculated by the formula

$m_{ij}' = (-1)^{i+j} D_{ij}$ (7), where D_{ij} is the determinant of the matrix obtained by deleting the i-th row and the j-th column of M.

It is well known ([7], Chapter 3) that the element n_{ij} of the fundamental matrix N gives the mean number of times in state S_i before the absorption, when the starting state of the AMC is S_j, where S_i and S_j are non absorbing states.

3. Ergodic Markov Chains

A MC is said to be an EMC, if it is possible to go between any two states, not necessarily in one step. It is well known ([7], Theorem 5.1.1) that, as the number of its steps tends to infinity (*long run*), an EMC tends to an *equilibrium situation*, in which the probability vector P_k takes a constant price $P = [p_1\ p_2\ ....\ p_n]$, called the *limiting probability vector* of the EMC. Therefore, as a direct consequence of equation (3), the equilibrium situation is characterized by the equation

$P = PA$ (8), with $p_1 + p_2 + ... + p_n = 1$.

The entries of P express the probabilities of the EMC to be in each of its states in the long run, or in other words the importance (gravity) of each state of the EMC.

Let us now demote with m_{ij} the mean number of times in state S_i between two successive occurrences of the state S_j, i, j = 1, 2, ..., n. It is well known then that $m_{ij} = \frac{p_i}{p_j}$ (9), where p_i and p_j are the corresponding limiting probabilities ([7], Theorem 6.2.3)

AN AMC MODEL FOR DECISION MAKING

DM is the process of choosing a solution between two or more alternatives, aiming to achieve the best possible results for a given problem. Obviously the above process has sense if, and only if, there exist more than one *feasible solutions* and a suitable criterion that helps the decision maker to choose the best among these solutions. It is recalled that a solution is characterized as feasible, if it satisfies all the restrictions imposed by the statement of the problem as well as all natural restrictions imposed onto the problem by the real system. For example, if x denotes the quantity of the stock of a product, it must be $x \geq 0$. The choice of the suitable criterion, especially when the results of DM are affected by random events, depends upon the desired goals of the decision maker; e.g. optimistic or conservative criterion, etc.

The rapid technological progress , the impressive development of the transport means, the globalization of our modern society, the enormous changes happened to the local and international economies and other relevant reasons led during the last 60-70 years to a continuously increasing complexity of the problems of our everyday life. As a result the DM process became in many cases a very difficult task, so that it is impossible to be based on the decision maker's experience, intuition and skills only, as it usually used to happen in the past. Thus, from the beginning of the 1950's a progressive development started of a systematic methodology for the DM process, which is based on Probability Theory, Statistics, Economics,

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10

DOI: http://doi.org/10.5281/zenodo.3662434

Psychology, etc. and it is termed as *Statistical Decision Theory (SDT)* [11].

1. The Steps of the DM Process

According to the nowadays existing standards the DM process involves the following steps:

- d_1: Analysis of the decision problem, i.e. understanding, simplifying and reformulating the problem in a way permitting the application of the principles of SDT on it.
- d_2: Collection from the real system and interpretation of all the necessary information related to the problem.
- d_3: Determination of all the alternative feasible solutions.
- d_4: Choice of the best solution in terms of the suitable (according to the decision maker's goals and targets) criterion.

One could add one more step to the DM process, the verification of the chosen decision according to the results obtained by applying it in practice. However, that step is extended to areas which due to their depth and importance for the administrative rationalism have become autonomous. Therefore, it is usually examined separately from the other steps of the DM process.

Note that the first three steps of the DM process are continuous in the sense that the completion of each one of them usually needs some time, during which the decision maker's reasoning is characterized by transitions between hierarchically neighbouring steps. Accordingly its flow diagram is represented in Figure 3 below:

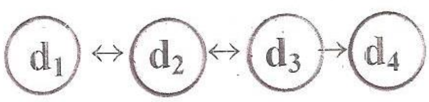

Figure 3: Flow diagram of the DM process

2. The Model

We introduce a finite MC having as states the steps d_i, i = 1, 2, 3, 4, of the DM process introduced in the previous section. Obviously d_1 is always the starting state. Further, we observe that, when the chain reaches the state d_4 (end of the DM process) it is impossible to leave it. This means that d_4 is the unique absorbing state of the chain. Therefore, since it is possible from any state to reach the absorbing state d_4 (see Figure 3), our MC is an AMC.

Taking into account the flow diagram of Figure 3 one finds that the transition matrix of the MC is

$$A = \begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ d_4 \end{array} \begin{array}{c} \begin{array}{cccc} d_1 & d_2 & d_3 & d_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ p_{21} & 0 & p_{23} & 0 \\ 0 & p_{32} & 0 & p_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

with $p_{21}+p_{23} = p_{32}+p_{34} = 1$.

Denote by $P_i = [p_1^{(i)}\ p_2^{(i)}\ p_3^{(i)}\ p_4^{(i)}]$ the probability vector of the MC, i = 0, 1, 2,.... . Then, since d_1 is always the starting state, we have that $P_0 = [1\ 0\ 0\ 0]$. Therefore, applying equation (3) one finds that

$$P_1 = P_0A = [0\ 1\ 0\ 0]$$
$$P_2 = P_1 A = [p_{21}\ 0\ p_{23}\ 0] \qquad (10)$$
$$P_3 = P_2 A = [0\ \ p_{21}+ p_{23}p_{32}\ \ 0\ \ p_{23}p_{34}]$$
$$P_4 = P_3A = [p_{21}^2+p_{21}p_{23}p_{32}\ \ 0\ \ p_{21}p_{23}+ p_{23}^2 p_{32}\ \ p_{23}p_{34}]$$

and so on.

We now bring the transition matrix A to its standard form A* and we make a partition of A* to sub-matrices as follows:

$$A^* = \begin{array}{c} \\ d_4 \\ d_1 \\ d_2 \\ d_3 \end{array} \begin{array}{c} \begin{array}{cccc} d_4 & d_1 & d_2 & d_3 \end{array} \\ \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & p_{21} & 0 & p_{23} \\ p_{34} & 0 & p_{32} & 0 \end{array}\right] \end{array}.$$

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10
DOI: http://doi.org/10.5281/zenodo.3662434

Then $Q = \begin{bmatrix} 0 & 1 & 0 \\ p_{21} & 0 & p_{23} \\ 0 & p_{32} & 0 \end{bmatrix}$ and applying equation

(6) one finds that

$$N=(I_3-Q)^{-1} = \frac{1}{p_{23}p_{34}} \begin{bmatrix} 1-p_{32}p_{23} & 1 & p_{23} \\ p_{21} & 1 & p_{23} \\ p_{21}p_{32} & p_{32} & p_{23} \end{bmatrix} = [n_{ij}],$$

i, j = 1,2,3. Therefore, since in our case d_1 is always the starting state, the mean number of steps taken before the absorption is given by

$$t = \sum_{i=1}^{3} n_{1i} = \frac{2+p_{23}p_{34}}{p_{23}p_{34}} \quad (11).$$

Obviously, the greater is the value of t, the more the difficulties that the decision maker faces during the DM process. In other words t provides an indication for the difficulty of the DM process. Another indication for the difficulty of the DM process is the time spent by the decision maker to complete the process, etc.

3. An Application

A company, say A, must decide about the proper place for building a new factory. The manager of the company wants to determine the probability for the DM process to be terminated in four steps and to estimate the mean number of steps needed before taking the decision. Here we analyze the DM process according to the previously presented AMC model:

d_1: Analysis of the DM problem

The analysis of the problem has shown that the profitability of the company's decision depends on the quality of the products of the existing in the area competitive companies.

d_2: Collection and interpretation of the necessary information

It turns out that there is only one competitive company in the area, say B, which produces three different products, say W_1, W_2 and W_3.

d_3: Determination of the feasible solutions

The funds available for the company A to build its new factory, as well as the already existing in the area factories and storehouses of the two companies A and B, suggest four favourable places , say P_1 , P_2 , P_3 and P_4 for the possible construction of the new factory. However, some additional information became necessary at this point in order to proceed to the choice of the best place.

$d_3 \rightarrow d_2$: Going back from d_3 to d_2

The market's research has shown that the expected net profits for the company A with respect to the favourable places for the construction of the new factory and the types of the products of the company B are those shown in Table 1

Table1: Expected net profits for the company A

$$\begin{array}{c|cccc} & P_1 & P_2 & P_3 & P_4 \\ \hline W_1 & 3 & 8 & 5 & 4 \\ W_2 & 4 & 2 & 6 & 5 \\ W_3 & 2 & 1 & 1 & -1 \end{array}$$

$d_2 \rightarrow d_3$: New transition from d_2 to d_3

From Table 1 it becomes evident that the feasible solution P_4 is worse than P_3 and therefore P_4 is rejected.

d_4: Choice of the best solution

The management of the company does not want to risk having low profits from the construction of its new factory, which means that it must adopt a conservative criterion for the choice of the best place for building it. Such a criterion that is frequently used is the *Wald's criterion*, which is based on the Murphy's law stating that the worst possible fact to be happen will happen. That criterion suggests to maximize the minimal possible for each case profits. In other words, since the minimal expected profit from the choice of P_1 is 2 monetary units and the minimal profit from the choice of P_2 and of P_3 is 1 monetary unit (see Table 1), according to the

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10

DOI: http://doi.org/10.5281/zenodo.3662434

Wald's criterion the place P_1 must be chosen for building the new factory.

From the above analysis of the DM process it becomes evident that $p_{21} = 0$ and $p_{23} = 1$. We also claim that $p_{32} = p_{34} = \frac{1}{2}$. In fact, when the MC reaches the state d_3 for first time, the probability of returning to d_2 in the next step is 1, since the collection and interpretation of new information is necessary. However, the second time that the MC reaches d_3 the probability of returning to d_2 in the next step is 0, since no more information is needed for the choice of the best solution. Therefore the transition probability p_{32} is equal to the mean value $\frac{0+1}{2}$ and $p_{34} = 1 - p_{32} = \frac{1}{2}$.

Replacing those values of the transition probabilities to the third of equations (10) one finds that $P_3 = [0 \;\; 0.5 \;\; 0 \;\; 0.5]$, i.e. $p_4^{(3)} = \frac{1}{2}$. Therefore, the probability for the DM process to be terminated in four steps is 50%.

Further, from equation (11) one obtains that

$$N = \frac{1}{0.5}\begin{bmatrix} 0.5 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0.5 & 1 \end{bmatrix}$$

Therefore $n_{11} = 1$ and $n_{12} = n_{13} = 2$. Thus, the mean number of steps for the completion of the DM process is t = 5 steps.

AN EMC MODEL FOR CBR

CBR is the process of solving problems based on the solutions of previously solved analogous problems (past cases). For example, a physician who cures a patient based on the therapy that has previously applied to patients presenting similar symptoms is using the CBR methodology. The use of computers enables the CBR systems to preserve a continuously increasing "library" of past cases and to retrieve in each case the suitable one for solving the corresponding new problem.

CBR first appeared in commercial systems in early 1990's and since then has been sued to create numerous applications in a wide range of domains including diagnosis, help-desk, assessment, decision support, design, etc. Organizations as diverse as IBM, VISA International, Volkswagen, British Airways, NASA, etc. have already made use of CBR in the above mentioned domains and in many more that are easily imaginable.

1. The Steps of the CBR Process

CBR has been formalized for purposes of computer and human reasoning as a four steps process involving the following actions:

- R_1: *Retrieve* the most similar to the new problem past case.
- R_2: *Reuse* the information and knowledge of the retrieved case for designing the solution of the new problem.
- R_3: *Revise* the proposed solution for use with the new problem. .
- R_4: *Retain* the part of this experience likely to be useful for future problem-solving.

Through the revision the solution is tested for success. If successful, the revised solution is directly retained in the CBR system's library; otherwise it is repaired and evaluated again. When the final result is a failure, the system tries to compare it to a previous analogous failure (transfer from R_3 back to R_1) and uses it in order to understand the present failure, which is finally retained in the library. When the CBR process is completed in R_4, it is assumed that a new analogous problem is forwarded to the system for solution. Therefore the process is transferred back to R_1 and a new circle is repeated. According to the above description the flow diagram of the CBR process can be graphically represented as shown in Figure 4.

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10
DOI: http://doi.org/10.5281/zenodo.3662434

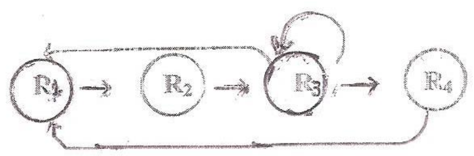

Figure 4: The flow diagram of the CBR process

For more details about the CBR process and methods we refer to [12] and to the relevant references included in that paper.

2. The Model

We introduce a finite MC having as states the four steps of the CBR process that have been described in the previous section. From the flow diagram of Figure 4 it becomes evident that in this case we have an EMC with transition matrix

$$A = \begin{array}{c} \\ R_1 \\ R_2 \\ R_3 \\ R_4 \end{array}\begin{array}{c} \begin{array}{cccc} R_1 & R_2 & R_3 & R_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ p_{31} & 0 & p_{33} & p_{34} \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array},$$

where $p_{31} + p_{33} + p_{34} = 1$.

Further, by equation (8) one finds that in the long run we have for the equilibrium situation of the EMC

$[p_1 \quad p_2 \quad p_3 \quad p_4] = [p_1 \quad p_2 \quad p_3 \quad p_4]$A or

$[p_1 \quad p_2 \quad p_3 \quad p_4] =$

$[p_3 p_{31} + p_4 \quad p_1 \quad p_2 + p_3 p_{33} \quad p_3 p_{34}]$

Consequently it turns out that

$p_1 = p_3 p_{31} + p_4$, $p_2 = p_1$, $p_3 = p_2 + p_3 p_{33}$,

$p_4 = p_3 p_{34}$ (12)

Adding by members the first three of the equations (12) one finds that

$$p_1 + p_2 + p_3 = p_3 p_{31} + p_4 + p_1 + p_2 + p_3 p_{33}$$
$$\Leftrightarrow p_3 = p_4 + p_3(p_{31} + p_{33})$$
$$\Leftrightarrow p_3 = p_4 + p_3(1 - p_{34}) \Leftrightarrow p_4 = p_3 p_{34}$$

Therefore, the fourth of the equations (12) is equivalent to the rest of them. Consider now the linear system L of the first three of the equations (12) and of the equation $p_1+p_2+p_3+p_4 = 1$. It is straightforward to check that the determinant of L is equal to

$$D = \begin{vmatrix} 1 & 0 & -p_{31} & -1 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & p_{33}-1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 4 - 3p_{33} - p_{31}$$

Also $$D_{p_1} = \begin{vmatrix} 0 & 0 & -p_{31} & -1 \\ 0 & -1 & 0 & 0 \\ 0 & 1 & p_{33}-1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 1 - p_{33}$$

Therefore, by the Cramer's rule one finds that

$$p_1 = \frac{D_{p_1}}{D} = \frac{1-p_{33}}{4-3p_{33}-p_{31}} = p_2. \quad (13) \qquad (13).$$

In the same way one also finds that

$$P_3 = \frac{1}{4-3p_{33}-p_{31}} \text{ and } P_4 = \frac{1-p_{33}-p_{31}}{4-3p_{33}-p_{31}} = \frac{p_{34}}{4-3p_{33}-p_{31}} \quad (14).$$

The values of the p_i's give the probabilities of the CBR process to be in step Ri in the long run, i = 1, 2, 3, 4, or in other words they give the importance (gravity) of each of the steps of the CBR process. Furthermore, since R_1 is the starting state of the EMC it becomes evident that the sum $m = m_{14} + m_{24} + m_{34}$ calculates the mean number of steps of the EMC between two successive occurrences of the state R_4. Therefore, the mean number of steps for the completion of the CBR process will be m+1, since it includes also the step R_4. With the help of equation (9) one finds that

$$m = \frac{p_1 + p_2 + p_3}{p_4} = \frac{1-p_4}{p_4} \quad (15)$$

ISSN: 0975 – 9808
IJINN (2020), 10(1):1-10

DOI: http://doi.org/10.5281/zenodo.3662434

It becomes evident that the greater is the value of m, the more are the difficulties during the CBR process. Another factor of those difficulties is the total time spent for the completion of the CBR process, which however is negligible when using computers.

EXAMPLE: A physician, in order to determine the disease and suggest the analogous treatment of a patient, takes into account the diagnosis and treatment of a previous patient having similar symptoms. If the initial treatment fails to improve the health of the patient, then the physician either revises the treatment (stay to R_3 for two successive phases), or gets a reminding of a previous similar failure and uses the failure case to improve the understanding of the present failure (transfer from R_3 to R_1).

Assume that the recorded statistical data show that the probabilities of a straightforward cure of the patient as well as of each of the above two reactions of the physician in case of failure of the initial treatment are equal to each other. Therefore $p_{31} = p_{33} = p_{34} = \frac{1}{3}$. Then equations (13) and (14) give that $p_1 = p_2 = \frac{1}{4}$, $p_3 = \frac{3}{8}$ and $p_4 = \frac{1}{8}$. That means that in this case the step of revision (R_3) has the greatest gravity among the steps of the CBR process.

Further, equation (15) gives that $m = 7$. Consequently the mean number of steps for the completion of the CBR process is 8.

4. Measuring the effectiveness of a CBR system

A CBR system should support a variety of retrieval mechanisms and allow them to be mixed when necessary. In addition, the system should be able to handle large case libraries with the retrieval time increasing with the number of cases.

Let us consider a CBR system including a library of n recorded past cases and let m_i be the outcome of equation (15) for case c_i, i=1,2,..., n. Each m_i can be stored in the system's library together with the corresponding case. Then we define the system's *effectiveness*, say E, to be the mean value of the m_i's of its stored cases, i.e. we have that $E = \frac{\sum_{i=1}^{n} m_i}{n}$ (16).

The more problems are solved through the given CBR system, the bigger becomes the number n of the stored cases in its library and therefore the value of E is changing. As n is increasing it is normally expected that E will decrease, because the values of the m_i's of the new stored cases will be normally decreasing. In fact, the bigger is n, the greater would be the probability for a new case to have minor differences with a past case, and therefore the less would be the difficulty of solving the corresponding problem via the CBR process. Thus we could say that a CBR system "*behaves well*" if, when n tends to infinity, then its effectiveness tends to 3, which, according to the flow-diagram of Figure 4, is equal to the minimum number of steps between two successive occurrences of R_4.

EXAMPLE: Consider a CBR system that has been designed in terms of *Schank's model of dynamic memory* for the representation of cases [13]. The basic idea of this model is to organize specific cases, which share similar properties, under a more general structure called a *generalized episode (GE)*. During the storing of a new case, when a feature of it matches a feature of an existing past case, a new GE is created. Hence the memory structure of the system is in fact dynamic, in the sense that similar parts of two case descriptions are dynamically generalized to a new GE and the cases are indexed under this GE by their different features.

In order to calculate the effectiveness of a system of this type we need first to calculate the effectiveness of each of the GE's contained in it. For example, assume that the given system contains a GE including three cases, say c_1, c_2 and c_3. Assume further that c_1 corresponds to a straightforward successful application of the CBR process, that c_2 is the case described in the

DOI: http://doi.org/10.5281/zenodo.3662434

example of the previous section, and that c_3 includes one "return" from R_3 to R_1 and two "stays" to R_3. Then m_1 = 3 and m_2 = 7. For calculating m_3 observe first that $p_{31} = p_{34} = \frac{1}{4}$ and $p_{33} = \frac{1}{2}$. Therefore, the second of equations (14) gives that $p_4 = \frac{1}{9}$ and equation (15) gives that m_3 = 8. Thus the effectiveness of this GE is equal to $E = \frac{3+7+8}{3}$ = 6. In the same way we calculate the effectiveness of all the other GE's of the CBR system and their mean value, which gives the system's total effectiveness.

Notice that a complex GE may contain some more specific GE's including some common cases (see Figure 3 in page 12 of [14]). Then we calculate the effectiveness of the complex GE by considering all its cases only once, regardless if they belong or not to one or more of the specific GE's contained in it.

An alternative approach for the representation of cases in a CBR system is the *category and exemplar model* applied first to the PROTOS system [15]. In this model the case memory is embedded in a network of categories, cases and index pointers. Each case is associated with a category. Finding a case in the case library that matches an input description is done by combining the features of the new problem into a pointer to the category that shares most of these features. A new case is stored in a category by searching for a matching case and by establishing the appropriate feature indices. The process of calculating the effectiveness of a system of such type is analogous to the process described in the previous example, the only difference being that one has to work with categories instead of GE's.

In a similar way one may calculate the effectiveness of a system corresponding to other case memory models including Rissland's and Ashley's HYPO system in which cases are grouped under a set of domain-specific dimensions [16], the MBR model of Stanfill & Waltz [17], designed for parallel computation rather than knowledge-based matching, etc.

CONCLUSION

The theory of MC's is one of the mathematical tools that are used in applications of AI characterized by randomness. In the present work we have modeled the decision making process in terms of an AMC on its phases and the CBR process by introducing an EMC on its steps. Further a measure has been obtained for assessing the effectiveness of a CBR system and examples have been presented to illustrate our results..

REFERENCES

1. Domingos, P., Lowd, D. (2009), Markov Logic: An Interface Layer for Artificial Intelligence, Morgan & Claypool Publishers, Williston, VT, USA.
2. Kemeny, J.G., Schleifer, A.J. & Snell, J.L (1962), Finite Mathematics with Business Applications, Prentice Hall, London.
3. Suppes, P. & Atkinson, R. C.(1960), Markov Learning Models for Multiperson Interactions, Stanford University Press, Stanford-California, USA.
4. Bartholomew, R.C. (1973), Stochastic Models for Social Processes, J. Wiley and Sons, London.
5. Kemeny, J. & Snell, J. l. (1976), Finite Markov Chains, Springer-Verlag, New York.
6. Zucchini, I.L., MacDonald, I.L., Langrock, R., Hidden Markov Models for Time Series, CRC Press, NY
7. Davis, M.H.A (2017), Markov Models and Optimization, Create Space Independent Publishing Platform, Columbia, SC., USA.
8. Voskoglou, M.Gr. (2017), Finite Markov Chain and Fuzzy Logic Assessment Models: Emerging Research and Opportunities, Create Space Independent Publishing Platform (Amazon), Columbia, SC, USA.
9. Voskoglou, M. Gr. & Perdikaris, S. C (1991)., A Markov chain model in problem- solving, International Journal of Mathematical Education in Science and. Technology, 22, 909-914.

DOI: http://doi.org/10.5281/zenodo.3662434

10. Morris, A.O. (1978), An Introduction to Linear Algebra, Van Nostrand Beinhold Company Ltd., Berkshire, England.
11. Berger, J.O. (1980), Statistical Decision Theory: Foundations, Concepts and Methods, Springer- Verlag, New York, 1980.
12. Voskoglou, M. Gr. & Salem, A-B. M, (2014), Analogy-Based and Case-Based Reasoning: Two Sides of the Same Coin, International Journal of Applications of Fuzzy Sets and Artificial Intelligence, 4, 5-51.
13. Schank, R. (1982), Dynamic memory; A theory of reminding and learning in computers and people, Cambridge University Press.
14. Aamodt, A. & Plaza, E. (1994), Case- Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, Artificial Intelligence Communications, 7(1), 39- 52.
15. Porter, B. & Bareiss, B. (1986), PROTOS: An experiment in knowledge acquisition for heuristic classification tasks, Proceedings of the 1st International Meeting on Advances in Learning (IMAL), pp. 159-174, Les Arcs, France.
16. Rissland, E. (1983), Examples in legal reasoning: Legal hypotheticals, *Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI),* Karlsruhe
17. Stanfill, C. & Waltz, D. (1988), The memory-based reasoning paradigm, Case-based reasoning Proceedings from a workshop, pp. 414-424, Morgan Kaufmann Publications, Clearwater Beach, Florida.